DOCUMENT RESUME

ED 076 606 TM 002 586

AUTHOR	Pyecha, John N.
TITLE	The Approach to the Design of the North Carolina Statewide Assessment of Educational Progress.
INSTITUTION	Research Triangle Inst., Durham, N.C.
PUB DATE	26 Feb 73
NOTE	16p.; Paper presented at Annual Meeting of American Educational Research Association (New Orleans, Louisiana, February 26, 1973)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Academic Achievement; Decision Making; *Educational Needs; *Educational Planning; Elementary Grades; Evaluation Criteria; Evaluation Methods; *Program Effectiveness; *Program Evaluation; Secondary Grades; *Statewide Planning; Technical Reports

ABSTRACT

The strategies and problems involved in designing the North Carolina Assessment Program are described. After being given the mission statement and continuing objectives of the North Carolina State Department of Public Instruction (NCDPI), the basic decisionmaking needs of State-level educational planners were identified by the NCDPI/Research Triangle Institute (RTI) design study team. Several alternative assessment plans were developed, costed, and evaluated. This interative design process led to the development of a statewide educational assessment program for evaluating the effectiveness of North Carolina's public elementary and secondary education programs, including a special emphasis on Title I programs. (Author)

FORM 850

PRINTED IN U.S.A.

FILMED FROM BEST AVAILABLE COPY

4.20

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

ED 076606

TM 002 586

The Approach to the Design of the North Carolina Statewide Assessment of Educational Progress

by

John N. Pyecha
Senior Educational Systems Analyst
Center for Educational Research and Evaluation
Research Triangle Institute
Research Triangle Park, North Carolina

Paper Presented at the
1973 Annual Meeting of the American
Educational Research Association,
New Orleans, Louisiana
26 February 1973

The Approach to the Design of The North Carolina Statewide Assessment of Educational Progress

I. INTRODUCTION

The State of North Carolina began the planning phase for the initial stage of a statewide assessment program in the early fall of 1970. The assessment plan was completed the following spring, a budget was prepared and funded in the summer of 1971, and data for the first assessment cycle were collected from a sample of approximately 12,000 sixth-grade students in the spring of 1972. These data included measures in four cognitive achievement areas (reading, mathematics, language arts, and career awareness) as well as in the areas of academic ability, student attitudes, and other noncognitive factors associated with achievement. The results were analyzed during the summer and fall of 1972. The dissemination plan was initiated in the winter of 1972. Implementation of subsequent stages of the assessment plan is pending state legislative action for additional funds to continue the program.

The purpose of this paper is to provide an overview of the North Carolina Assessment Program and to present a brief description of the approach and problems involved in designing it. The other papers presented during this AERA discussion session will focus on the methodology of certain key components of the program; i.e., the instrumentation package, the data collection plan, the complex sample design, the analysis plan, and the dissemination and utilization plan.

An overview of the development and implementation of the initial phase of the North Carolina Assessment Program is presented in Section II of this paper. The planning rationale and scenario are described in Section III. These preliminary discussions provide a background for the planning approach presented in Section IV.

II. OVERVIEW OF THE DEVELOPMENT AND IMPLEMENTATION OF THE INITIAL STAGE OF THE NORTH CAROLINA ASSESSMENT PROGRAM[1/]

The need for a statewide assessment program evolved through the expressed interest of various groups concerned with improving education in North Carolina. Information about the status of educational progress in the public schools is needed by citizen groups, legislators, local school boards, and those with responsibility for educational leadership at the state level as they all work for the improvement of education in North Carolina.

The State Superintendent, aware of the need for more educational performance information, initiated the state assessment as part of a total effort in better management practices. A planning grant relating to state assessment of education was approved by the State Board of Education in 1970. The grant was made to the Research Triangle Institute (RTI), a North Carolina not-for-profit research firm with extensive experience in the National Assessment of Educational Progress project.

RTI, in this planning contract, was charged with assisting the State Department of Public Instruction in designing a statewide assessment program that would generally result in (1) more accurately identifying the educational needs and priorities within the state; (2) increasing the general public's understanding of the attainments, needs, and problems of the schools; and (3) supplying more objective information which can be used by the legislature as it considers the educational needs of the state. In addition, the design for the initial phase of the program was to include a special emphasis on assessing the educational status of students in special programs supported and administered with ESEA Title I funds.

The RTI planning study involved the collection of information from other states on their "assessment successes and problems," as well as the evaluation of various approaches and instruments available for assessing educational progress. Inputs and reactions to the design for a state assessment were sought from all divisions within the State Education Agency. Furthermore,

---

1/ Most of this section has been excerpted from: The 1971-72 State Assessment of Educational Progress in North Carolina. Raleigh, N.C.: State Department of Public Instruction, December 1972, pp. 2-6.

several intensive preassessment briefings were held with local superintendents in the spring and summer of 1971. Many educators, including classroom teachers, had input into the assessment plan developed for North Carolina.

During the months of November and December (1971), RTI collected data from the State Education Agency and from other agencies of state government in order to select probability samples which would lead to reliable estimates of sixth-grade student achievements for: (1) the state as a whole; (2) the three geographic regions of the state; (3) the three types of communities most prevalent in North Carolina; (4) whites and non-whites; (5) males and females; and (6) the educationally disadvantaged children served by ESEA Title I.

By the end of December (1971), the schools in the sample had been selected and letters were sent to superintendents to provide information on the assessment project as well as to define the project's needs from his unit. A team of 15 consultants from several divisions of the State Education Agency was formed to be of assistance in coordinating the collection of information from local school units. With the support of the 130 superintendents whose units were chosen for the sample, the assessment project received 100% cooperation from the 448 schools which had been randomly selected.

After careful study and deliberation, the State Education Agency selected the assessment instruments and the method through which they would be administered. The total testing time per child was limited to one day by assigning individual students to take only one of four assessment batteries; i.e., a form of matrix sampling was used. The assessment batteries were in four cognitive areas--reading, language arts, mathematics, and career awareness. In addition, an assessment was made of academic ability and selected student and school background factors which have been shown to be associated with achievement.

The cognitive assessment instruments were reviewed by a group of school psychologists and two LEA Superintendents who met in October. With their concurrence, the final selection of cognitive and academic ability tests was made in December. Reading, language arts, and mathematics were measured by the Iowa Test of Basic Skills; career awareness was measured by the Cognitive Vocational Maturity Test, an instrument that was developed at

the Center for Occupational Education at North Carolina State University; and academic ability was measured by the Lorge-Thorndike Intelligence Test. Due to the scarcity of noncognitive instruments with established validity, the Division of Research developed an instrument to explore students' perceptions in several noncognitive areas reportedly associated with academic achievement; e.g., self concept and attitude toward school. The Divisions of Language Arts, Mathematics, and Compensatory Education were consulted about special measurements of the schools' educational program. The Mathematics Division developed a supplementary test for Mathematics computation. Field tests of the assessment instruments were undertaken in February (1972).

Local school districts aided in the assessment program. The principals provided rosters of their sixth graders for sampling purposes and the superintendents appointed a total of 130 assessment coordinators. These assessment coordinators then scheduled personnel from the local superintendents' offices to do the test administrations. Personnel from the local superintendents' offices were used rather than classroom teachers in order to insure greater standardization and control of the testing situation. Approximately two hundred and fifty of these professionals averaged about five days working on the testing. A series of training sessions was conducted in mid-April (1972). Each test administrator spent one day at the training session where materials and precoded answer sheets were distributed and reviewed. Questions on the activities and responsibilities of each test administrator were cleared up at these sessions. A special telephone number at the State Education Agency ("Assessment Hotline") was established to provide assistance in handling unforeseen situations.

The typical school's involvement in the assessment program was limited to only one day. During this day of assessment, approximately twenty sixth-grade students who had been selected on a probability basis took an assessment battery. School information was obtained from the principal or his designate by having him complete a brief school questionnaire. The homeroom teacher for each of the students in the sample was asked to complete a short student questionnaire.

The students recorded their answers on machine-scannable forms. The test administrators checked these sheets for accuracy and coded the student

information onto that same scannable form. This coding and marking process was reviewed by the Assessment Coordinator who mailed his unit's materials to RTI where they underwent a third editing. After all editing questions were resolved, the sheets were sent to the Measurement Research Center (MRC), one of the nation's largest and most respected test scoring companies, for scanning and scoring. An important part of MRC's processing procedure is the machine edit conducted by the computer on all sheets before the students' scores are determined. MRC also combined the school information that was collected from the State Education Agency and from the principal's questionnaire with the students' responses and scores. The merged information was placed onto one large reel of computer tape. When these tape files were received by RTI, they were subjected to final edits that led to the deletion of 5-10 records (out of approximately 11,000) that were unusable for a variety of reasons.

Preliminary tallies and univariate analyses of assessment data were supplemented with some linear regressions in order to provide some insights into the relative relationships among the major variables of the study. The data analysis phase, completed by RTI in mid-October (1972) produced "region" and "type-of-community" student norms, as well as some initial analyses of the major variables in the study. Additional work was completed by the Division of Research as they reviewed the RTI technical report and prepared a highlights report for the State Board of Education. The release of data is being coordinated through the Office of the Assistant Superintendent for Research and Development.

III. PLANNING RATIONALE AND SCENARIO

A. Rationale

The decision-making rationale as presented by Stufflebeam, et al.,[2/] as a basis for evaluation also provided the basis for the approach used by RTI to assist North Carolina in planning and designing the assessment program. That is, a decision maker must be aware of several options or alternative assessment approaches and have a basis for choosing among them. The basis for making these selections is derived from the personal or organizational values of the decision maker, which in turn reflect the time, resource (funds and personnel), and political constraints under which the program must operate.

Following this rationale, RTI delineated the assessment options for the State Education Agency (the decision maker) and provided the Agency with the necessary information for differentiating or ordering them. Given this information about each option, the Agency (through the Division of Research) selected the option that best conformed to its values. As discussed in subsequent pages of this report, this rationale was applied in an iterative fashion that involved the presentation of many options and required a series of interrelated decisions.

B. Scenario

The most effective way to begin designing an assessment program would be to start with a list of detailed objectives or specifications that serves to define the purposes of the desired program (who needs what kind of information to make which decisions under what circumstances) and a good understanding of the personal and organizational values of the decision maker. Unfortunately, however, one can reasonably expect the planning process to not proceed in this fashion.

2/ Daniel L. Stufflebeam, et al. Educational Evaluation and Decision Making. Itasca, Ill.: F. E. Peacock Publishers, Inc., 1971, pp. 38-44.

For example, defining the purposes of the desired program is complicated by severe limitations in our current understanding of the decision-making process and of the methodologies for relating statewide assessment results to decision making. Issues in this area center around such problems as:

1) Trying to understand and explain the decision-making process through the use of models that assume the process to be a rational one, an assumption that appears to be tenuous with respect to actual decision-making situations.
2) Developing an adequate taxonomy of educational decisions.
3) Linking the assessment program to the decision makers it will ultimately serve.
4) Untangling the complex interrelationships between the various decisions being made by the users of state assessment results.[3/]

In addition, one can reasonably expect some misunderstanding among state education agencies as to the resources required to implement and maintain a comprehensive assessment program. Few states have had the "assessment experience" that leads to an awareness and appreciation of the scope and magnitude of such a program, including its many technical and operational problems and the interpretive limitations of its output; e.g., the inability of survey data to show cause-effect relationships. Compensating for this lack of experience by sharing assessment program cost data from other state or local educational agencies can be extremely misleading since current accounting procedures rarely provide accurate tabulations of incurred expenses (man-hours expended, overhead, travel and per diem, communications, reproduction, etc.) that might have been spread out over a number of agencies and/or departments.

As a result the RTI/North Carolina planning endeavor began with a list of general objectives for the desired assessment program, a mission statement and list of continuing objectives of the Department of Public Instruction and a rather optimistic expectation on the part of the State Education Agency as to the scope of the program that a given level of resources can support. The planning approach, therefore, was designed to ferret out

---

3/ These and other problems related to conceptualizing the decision-making process are presented in the reference in footnote 2; i.e., Daniel L. Stufflebeam, et al., pp. 16-18.

critical "purpose" information and to illuminate the resource demands, problems, and limitations associated with various types of state assessment programs. The greater the degree to which these tasks can be completed during the early stages of the planning process, the greater the degree of effectiveness to be attained in the subsequent planning stages.

It should be noted that the above scenario for the North Carolina planning study is not atypical based on subsequent RTI experiences in assisting four other state educational agencies in planning and/or conducting state assessment programs.

IV. DESIGN APPROACH

The planning study was conducted in close liaison with personnel from four divisions within the State Department of Public Instruction, i.e., the Divisions of Research, Program Services, Development, and Compensatory Education. The Division of Research served as the primary coordinating agency for obtaining inputs and feedback from the other divisions within the department, as well as from representative groups of superintendents and classroom teachers from school districts throughout the state.

An iterative planning approach was used whereby alternative assessment plans were developed, costed, and presented to these groups for consideration and feedback. Each presentation of an alternative plan(s) was accompanied by a written or verbal discussion of the strong and weak points of its (their) key features. On the basis of this feedback, a new alternative, or set of alternatives, was developed and the iteration or planning stage was repeated.

This iterative process was continued, within the time and funds of the planning contract, until an assessment plan was formulated that appeared to best meet the state's educational planning needs within limitations of actual or projected resources (funds and personnel), time, and political constraints for implementing and supporting the program.

Each alternative plan or iteration was presented and evaluated in terms of six components or tasks that were identified as being essential to the design of an effective statewide assessment program:

1) Management and Staffing.
2) Instrumentation Selection and/or Development.
3) Sampling Design.
4) Data Collection and Processing.
5) Data Analysis.
6) Reporting and Dissemination.

Some of the questions that had to be explored and resolved with respect to each of these six assessment components are listed below. Answers to each of these questions represent alternative assessment approaches or strategies.

1. Management and Staffing

   Is the assessment program to be a "one-shot-affair" or a continuing and ongoing program?

   What in-house staff capabilities does the State Education Agency have?

   Are funds available for contracting assessment-related services that cannot be performed by the available in-house staff?

   Does the State Education Agency plan to develop full in-house capability for managing and operating the assessment program?

   If so, over what period of time will this in-house staff capability be developed? If not, for which tasks or subtasks does the State plan to develop an in-house capability?

2. Instrument Selection and/or Development

   What student grade and/or age levels will the assessment program cover?

   What are the cognitive, affective, and/or psychomotor areas to be assessed?

   Should the assessment instruments be criterion-referenced or norm referenced?

   If norm-referenced, which tests will be used?

   If criterion-referenced, does the State plan to develop its own behavioral objectives, or use existing behavioral objectives?

   If criterion-referenced, will the test items come from existing item pools, or will they be especially created, or both?

   If the total time required to administer all student instrumentation exceeds a desirable time limit per student, will matrix sampling be used?

   If matrix sampling is used, how will instrumentation items be placed in separate packages while taking into consideration such factors as item format, placement location, mode of administration, subject area coverage, administration time, and item degree of difficulty?

   Will the norm- or criterion-referenced item types be machine scoreable?

Will the State set standards of performance in advance of conducting the assessment?

If so, how will such standards be set and what population subgroups will be involved (e.g., parents, students, educators, and other tax-paying citizens)?

What student, school, and/or teacher background information is desired?

Can the required student, school, and/or teacher background information be retrieved from existing data bases?

If not, will it be necessary to develop questionnaires or interview techniques for collecting these data?

3. Sampling Design

What level of statistical precision is required of the output estimates to be reported for the State population and/or for the various State subpopulations of interest?

What will constitute the units of the first-stage sampling frame (e.g., counties, school districts, schools)?

How will this first stage be stratified (e.g., enrollment size, size of community, or socioeconomic data for relevant population)?

How will first-stage units be allocated to the strata?

How will first-stage units be selected within the designated strata (e.g., with equal probabilities or with probabilities proportional to size of target population)?

What will constitute the units of the sample frames for subsequent stages in the sampling procedures; how will these units be stratified in each stage; and how will they be selected within the designated strata of each stage?

4. Data Collection and Processing

Will assessment items be administered by State, school, or contracted personnel, or by some combination of the three?

Are test items to be administered to groups of students, individual students, or both?

Will test administration instructions be taped?

Will open-ended exercises be used?

Will background information pertinent to students, programs, teachers, and schools be obtained through actual site visits by survey teams, or through mail surveys?

If background information is to be gathered by survey teams, will these teams consist of State or of contracted personnel, or some combination of the two?

When will the assessment data be collected?

Will the completed instrumentation be optically scanned?

What edit checks will be performed and how will any errors be resolved?

Will the assessment data be stored on magnetic tape files?

5. Data Analysis

Will the analysis be primarily descriptive in orientation, i.e., generally limited to the calculation of means, standard deviations, and percentile distributions for the various dependent variables within specific subgroups of the state population as defined by combinations of appropriate dependent variables?

Should the analysis be extended to include an investigation of the relationships between selected dependent and independent variables for various subgroups of the State population?

Does the State want to pursue the establishment of prediction equations for student achievement levels in terms of selected school, teacher, and student background variables?

6. Reporting and Dissemination

What reporting variables and reporting groups are desired for the assessment program?4/

What types of reports are to be prepared?

How will these reports be disseminated?

What additional dissemination activities are to be employed to insure maximum utilization of assessment results (e.g., workshops for teachers and curriculum specialists)?

Will audio and slide and/or TV presentations be developed?

What preassessment strategies are to be employed to insure that program goals, aims, purposes, and procedures are fully understood?

4/ A reporting variable is defined as a primary characteristic or set of characteristics that serve(s) to define the group of students for which information is desired and for which output measures are to be reported. Each reporting variable has two or more reporting groups; e.g., sex is a reporting variable, males and females are reporting groups. Examples of other reporting variables include: (1) such student-related variables as sex, SES, race, longevity in the state's educational system, and the type and degree of participation in extra-curricular activities; (2) such school-related variables as public and/or nonpublic status, type and size of community in which school is located, size of enrollment, teacher/student ratio, per pupil expenditure levels, regional location, and selected characteristics of staff and student body; and (3) process-related variables pertinent to instructional approach and/or program participation.

The initial planning iteration was designed to provide an overview of all six of these components. The range of alternatives for each component was presented and discussed in a general manner so as to project the "big assessment picture"--that is, to convey the general magnitude and complexity of a comprehensive assessment program and to provide a basic understanding of the interrelationships between the six components and how a choice of an alternative in one component might effect the selection of alternatives in one or more of the other components.

It was especially important to introduce the reporting and dissemination component during the initial planning iteration(s). By exploring various types of outputs that might be generated through a state assessment program, State personnel were better prepared to focus in on the purposes of the assessment--to really consider whether or not specific types of information would be meaningful to certain key decision makers. (Preparing table shells and/or presenting examples of results from other state assessments, though not used in planning the North Carolina program, were extremely helpful strategies in this regard during subsequent RTI planning studies.) It is easy to omit this important component, or to treat it lightly, in the enthusiasm created by a desire to "get something started." Every one involved in the planning process has a perception of what the program will provide and, quite often, these perceptions vary considerably. It is wise, therefore, to draw these perceptions out "early in the game" and, if necessary, bring them into a reasonable focus.

In addition to painting the "big picture," the initial planning iteration must generate sufficient feedback to provide direction for the next iteration. For example, the range of available and projected resources might be narrowed; or an instrumentation decision might be made on the normative versus criterion-referenced approach; or the grade and/or age levels to be assessed might be better defined. One can not hope to explore fully all possible alternative designs; hence, State personnel must narrow the range of feasible assessment alternatives by disgarding those alternatives that appear impractical or undesirable as soon as they can be identified. Several such alternatives were identified by North Carolina during the first planning session.

As the planning process proceeded through successive iterations, distinctions between alternatives became finer and the costing became more exact. That is, the planning tended to move from the "general" to the "specific." Approximately 20 planning iterations were required to arrive at the final North Carolina plan. (It is important to note that the planning process never really stops--it goes on during operational stages as one tries to implement the "final" plan.)

Because of the interrelationships between alternatives, it is difficult to establish a distinct sequence by which the various alternatives were selected during the course of these iterations. For example, management and staffing decisions depend upon the nature of the assessment tasks, which in turn depends upon decisions in all five of the other six major components; or, a sample design decision might depend upon how the results are to be reported, or on the type of instrumentation to be used, or on how the data are to be collected. As a result, the process was not simply one of selecting a final alternative in each of the six major components, one component at a time. Generally speaking, each iteration tended to "shave" a little off of each component as alternatives were cast aside. It was essential, however, that the target population be defined and the instrumentation and resource availability questions be resolved during the early iterations.

A key factor in the successful application of this planning process in the North Carolina study was the decision-making capabilities of involved State Education Agency personnel. Some Agency decisions pertinent to the selection of alternatives were appropriately based on consultation with advisory groups, persons in higher authority, and consultants in various divisions. However, it was often necessary to make many decisions on the spot; otherwise, the process could have extended indefinitely. Personnel from the State Education Agency performed this function well. When provided with supporting information at each point in the decision tree, they effectively isolated the pertinent elements and quickly arrived at firm and clear-cut decisions.

In conclusion, the design and successful implementation of the North Carolina assessment program would not have been possible without the fine collaborative effort of the total North Carolina educational community--students, classroom teachers, local superintendents, all divisions within the State

Education Agency, the administration within the State Education Agency, and the Research Triangle Institute. The positive relationships between these research and educational agencies demonstrated the value of an open and cooperative exchange of ideas, experience, and services.